Computer-aided Learning Validation: A CAI-Critical Mission

Thomas J. Handler, MD, Patrick Lynch, MS, and C. Carl Jaffe, MD
The Center for Advanced Instructional Media and the Center for Medical Informatics
Yale University School of Medicine, New Haven, Connecticut

Despite a marked increase in computer-assisted instructional applications (CAI) over the past few years, little attention has been paid to revising traditional approaches toward educational testing. This paper reports a CAI project that emphasizes integrating the testing and training of visual judgmental capacities of health care professionals. It takes advantage of the computer's ability to display digital video segments and to record and compare user learning accomplishment and at the same time a normative performance scale can be developed. The program uses a method which, in addition to validating the efficacy of the project itself, collects data and stratifies users' level of proficiency by integrating pre-test and post-test modules. Routine incorporation of these principles in CAI may provide a more effective means of correctly evaluating the individual's mastery of a topic.

INTRODUCTION

The potential of computers to assist in education has been recognized since their early development. Initially, computer-aided instruction (CAI) required costly main-frame computers and specialized knowledge of a programming language to build specific applications. These requirements tended to hamper effective CAI development by compartmentalizing the construction of CAI content and making incremental improvements and didactic experiments difficult. With the arrival of powerful, less costly personal computers, and more natural, language-like authoring programs, exploration of new experimental didactic paradigms has been possible. Early concern that CAI would be inferior to traditional teaching methods has been countered by a number of studies, which demonstrate that carefully focused CAI can be at least as effective as standard teaching methods for appropriately chosen subject matter.(1, 2). One area of continued concern is project evaluation, since there have been few articles that address the efficacy of specific CAI projects or permit a general approach to validating didactic effectiveness(3-5).

This paper will show an integrated mechanism for combining didactic and evaluative processes within a CAI project. We demonstrate an approach that goes beyond merely providing expository teaching and addresses the critical issues of feedback mechanisms and evaluation of end-user performance.

BACKGROUND

Echocardiography (ECHO) and equilibrium radionuclide angiocardiography (ERNA) are common clinical imaging tests used to evaluate cardiac function. One component of this evaluation is the estimation of left ventricular ejection fraction (EF). In the ERNA method, a portion of the patient's red cells are radioactively labeled, injected intravenously, and imaged with a gamma camera gated to the electrocardiogram. By obtaining computer calculations of the maximum radioactive counts over the left ventricular region in systole and diastole, the EF can be estimated. The reproducibility of the radionuclide method has been reported to be as high as 95%.(6, 7). ECHO uses acoustic waves to produce real-time images of the heart. Computerized quantitative methods are less reliable on the ultrasound images than on ERNA because of substantial signal clutter. Visual estimation of the EF using ECHO has been promoted but is not widely used because it requires prolonged training of visual judgment.(8) Also, the feedback mechanisms required for such training are not commonly available in the clinical setting.

OBJECTIVE

Our goal was to design a CAI project to teach visual estimation of left ventricular ejection fraction on echocardiography by computer display of digital video sequences integrated with an efficient immediate feedback technique. The target users were radiology residents, ultrasound technologists, and cardiology fellows. We collected data to evaluate the didactic efficacy of the project, and in addition, we devised an analytic methodology to stratify the level of a student's knowledge while developing a normative standard.

DESCRIPTION OF THE PROJECT

We developed "The Functional Heart: Echocardiography of Ventricular Contraction" using Supercard (Allegiant Corp., San Diego, Calif.), a multimedia authoring tool for the Macintosh computer. This natural language-like software was chosen because it lent itself to group authoring and to easy revision by the principal author-educators.

The first component of the CAI application offers the user a progressive, linear, didactic exposition of the process and of the visual techniques used for evaluation of cardiac ejection fraction. The bulk of the application is a series of 120 clinical cases displayed as short digital video clips of the cardiac acoustic image from two or three orthogonal views. Cases were included if the patient had both an ERNA and an ECHO within 24 hours and if the ECHO images were of high quality. The cases were divided into 7 major functional groups ranging from poor to good cardiac function as determined by ERNA EF. At least 10 different cases were included within each group.

Cases can be called up by the user in one of two ways. In the "learning mode", the user selects the desired range of EF and is shown a series of cases along with the calculated ERNA EF value. (Fig. 1) In the "testing mode", cases are presented in random order and the user is required to estimate the EF before instant feedback is provided. (Fig. 2) For the purposes of the training process, the ERNA EF is considered the "true value" and is provided as feedback to reinforce the user's visual judgment of the function displayed in the echosequence.

A stand-alone evaluation module composed of sixty ECHO cases was also developed for pre-test and post-test use. At logon, users were asked to record a self assessment of experience and confidence in echocardiography before they could proceed with viewing cases. After examining each case, the user typed the visually estimated cardiac ejection fraction into the computer keyboard and the response was logged to a separate file for later statistical analysis and creation of a normative database.

METHOD

The subjects, all of whom had different levels of echocardiographic experience, were technologists, residents and fellows rotating through the cardiac echo section. Each subject used the evaluation module described above. Three experienced clinical echocardiographers also participated. Scatter plots of the visually estimated ECHO EFs and ERNA EFs were drawn from the data acquired. Linear regression and correlation coefficients were calculated using JMP statistical software package (SAS Institute Inc., Cary, N.C.) (Fig. 3).

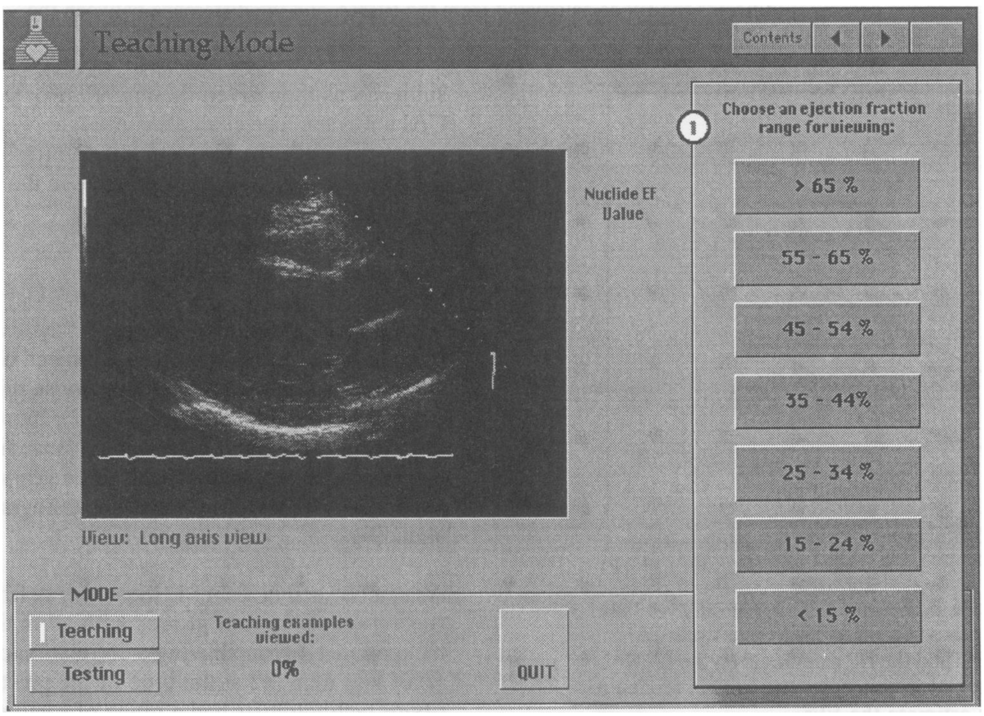

Figure 1. Demonstration of the teaching mode of the Functional Heart Project. The user can select the group of cases he or she is interested in seeing by pointing and clicking on one of the buttons at the right. The computer randomly chooses a case to be shown within the range of EFs selected. As the images are presented, the ERNA calculated EF is shown to the right of the ECHO image thus reinforcing the correlation between the image and the EF.

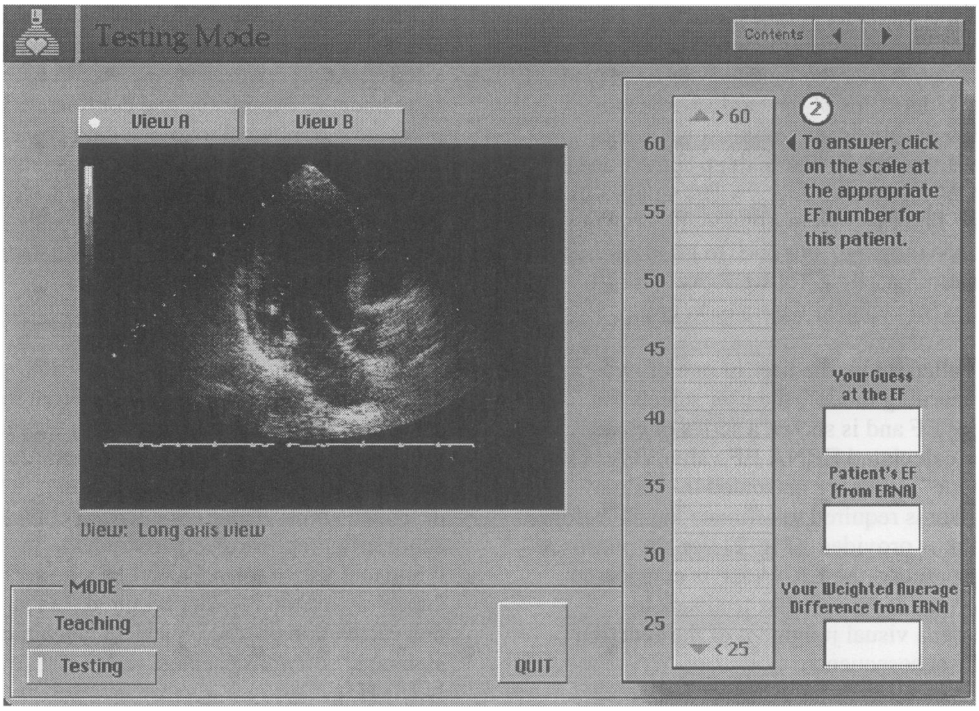

Figure 2. Sample page from the testing mode of the Functional heart project. The computer randomly selects a case and the user estimates the EF by clicking on the scale box on the right. Then the computer indicates the EF as calculated from the ERNA and the difference between the two values.

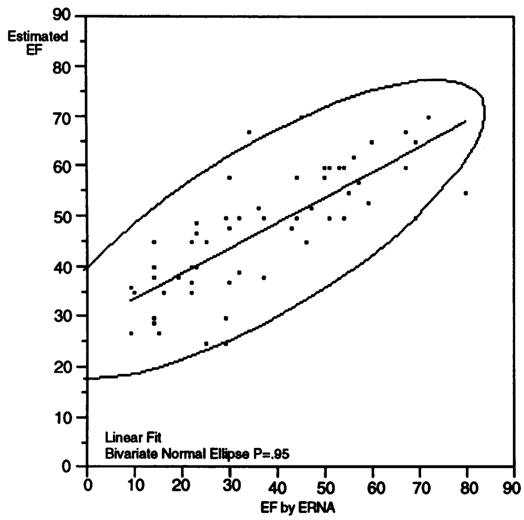

Figure 3. Example of a scatter plot that compares the EF as estimated from the ECHO and the EF calculated on the ERNA. Note that in this case, the subject tended to over estimate the EFs.

Those individuals who were not experienced in ECHO participated in a paired pre- and post-test study to assess the efficacy of the main CAI project. Each was asked to use the evaluation module prior to using the master CAI software. During the use of this pre-test module, no feedback was provided to the user. The subject was then given the opportunity to use the CAI software. The evaluation module was repeated as a post-test once the subjects felt comfortable with the feedback-provided material covered in the main CAI application.

RESULTS

A total of 25 individuals completed the evaluation module of the Functional Heart Project. Subjects were divided into several groups: those with no experience in echocardiography (12); those with at most a few months of experience (4); senior ultrasonographers with at least three years of experience (6); and the three echocardiography attendings.

As can be seen in Table 1, there was definite separation of the four groups with all in the inexperienced group having a correlation coefficient (r) of less than .82 at the time of the pre-test. The senior technologists had a range of r values from .83-.91, and the attendings had r values greater than .92.

Seven radiology residents, who had no prior experience with echocardiography, completed the paired pre- and post-test study. On average, the previously inexperienced cohort spent 2 1/2 to 3 hours using the main CAI application (range 2 to 6

SUBJECTS	r=
1	0.716
2	0.748
3	0.676
4	0.672
5	0.787
6	0.778
7	0.816
8	0.781
9	0.676
10	0.811
11	0.747
12	0.740
13	0.863
14	0.840
15	0.828
16	0.863
17	0.864
18	0.825
19	0.852
20	0.844
21	0.903
22	0.913
Attending 1	0.919
Attending 2	0.923
Attending 3	0.941

Table 1. Correlation coefficients(r) for cardiac ejection fraction that compares the EFs estimated on ECHO with EFs calculated on ERNA. Each subject was shown 60 cases. Subjects 1-12 had little or no experience with echocardiagraphy. Subjects 13-16 were cardiology fellows who had spent some time on the ECHO service prior to using the evaluation module. Subjects 17-22 were senior technologists who had at least three years of experience on the echocardiagraphy service.

hours). The analysis of the data obtained in this study is shown in table 2. All of the subjects showed significant improvement (p=.02) on the post-CAI evaluation as determined by the Wilcoxon signed-rank test.

DISCUSSION

Typically, students are evaluated, at least in part, by examination performances based on the number of correct versus incorrect answers. These black and white measurements are inadequate for assessment of competence in medical education, particularly in situations requiring clinical judgment. It is reasonable to expect senior residents to know more than junior residents, yet this difference is difficult to quantify based on the typical tests. In addition, some required medical knowledge does not conveniently fall into clearly definable categories of right or wrong. For carefully chosen learning tasks that lend themselves to quantitative results, the use of scatter plots and correlation coefficients—such as those used in evaluating the efficacy of this CAI project—can provide a useful solution. It also gives rise to the development of a normative performance pattern based on the visual judgment required for the given task.

SUBJECTS	Pre-CAI	Post-CAI
1	0.716	0.841
2	0.748	0.861
3	0.676	0.787
4	0.672	0.770
5	0.787	0.833
6	0.778	0.842
7	0.815	0.814

Table 2. Comparison of the correlation coefficients of visual estimation of EF on ECHO compared to ERNA calculated EF before and after the use of the Functional Heart project for seven radiology residents with no echocardiography experience prior to using the CAI project. Each estimated the cardiac ejection fraction on 60 cases pre- and post use of the project. There was significant improvement in the post test correlation coefficient as determined by the Wilcoxon signed-rank test (p=0.02).

The division between experienced and inexperienced individuals is clearly demonstrated in Table 1. Further stratification into categories of untrained, trained and expert individuals is possible if certain reasonable assumptions are made. If the senior technologists are considered to represent a trained group, then their correlation coefficients can be used as a yardstick for evaluating the performance of others. Any individual with an r less than .82 is considered untrained in ECHO; an r greater than .83 indicates adequate training; and anyone with an r greater than .92 can be considered an expert having achieved a similar score to the attendings in this study.

The training implication of this measurement should be clear. If a resident or fellow scheduled to rotate on the ECHO service, is shown to have an r value less than .82, he or she will likely benefit from training using the CAI project. If that individual has already received training and scores less than .82 it indicates the need for remedial work. In another scenario, if an attending's qualifications are being reviewed for consideration of coverage of the ECHO service, the level of expertise can be partially validated by the correlation coefficient.

As mentioned previously, the testing/feedback method initially was designed to evaluate the efficacy of CAI project. By documenting that the correlation coefficients can be stratified according to the level of experience we have shown that the ECHO images have sufficient information to permit reliable and reproducible diagnostic evaluation. More importantly, we have shown that the use of this CAI application produces a statistically significant improvement in visual judgment.

Our evaluation method has the added benefit of providing data that can be used to provide instructional feedback to the user. By creating a scatter plot of the visually estimated EFs versus the ERNA EFs and plotting the regression line, one can provide quick, graphic feedback to the user. His or her tendency to over or underestimate the EF at one or both ends of the scale can be seen, as shown in Figure 3. This information should be helpful in improving visual judgment.

We have demonstrated a method that not only provides a means of evaluating the effectiveness of a CAI project, but also provides a means of stratifying levels of proficiency. This particular project was made easier because EFs are numerical values that allow for standard statistical analysis. Where the subject matter does not permit a numerical value or even where there may not be a known "right answer", the method can be modified to examine the reproducibility of visual decisions and still be useful. In that situation the method would evaluate intra-observer reproducibility. Demonstration of consistent performance implies that the subject is performing a decision-making task based on consistent internalized criteria, although no judgment can be made as to whether the criteria are right or wrong.

FUTURE DIRECTION

Economic constraints on medical training will likely necessitate more cost-effective means of educating medical personnel. Although not specifically examined in this study, we believe that there is a point at which further time spent using the CAI project does not result in meaningful improvement in level of training. We plan to test this hypothesis in our next project as it may provide a quantitative means to determine the length of time needed to master a particular skill. This could have an impact on the curriculum of training programs.

In the coming managed care environment, there will likely be a need to document quality control and to certify level of training of providers. For example, the FDA currently requires certification of mammography centers and might expand this to include the mammographers, especially in light of Elmore's recent paper that demonstrated moderate inter- and intra-observer variability amongst radiologists performing mammography.(9). It is not unreasonable to expect similar findings and implications in other areas of medicine. The technique described in this paper can be applied to more than echocardiography and can provide a straight forward and reproducible means of documenting a physician's level of expertise.

Acknowledgments

We'd like to thank Richard Shiffman, Prakash Nadkarni, and Matt Healy for their assistance in writing this paper.

References

1. Jelovsek FR, Adenbonojo L. Learning principles as applied to computer-assisted instruction. MD Comput 1993;10(3):165-172.
2. Kulik JA, Kulik CL, Cohen PA . Effectiveness of computer-based college teaching: A meta-analysis of findings. Rev. Educ. Res. 1980;4:525-535.
3. Lyon HC, Healy JC, Bell JR, et. al. PlanAlyzer, an interactive computer-assisted program to teach clinical problem solving in diagnosing anemia and coronary artery disease. Academic Medicine 1992;67(12):821-828.
4. Lonwe B, Heijl A. Computer-assisted instruction in emergency ophthalmological care. Acta Ophthalmol 1993;71:289-95.
5. Chew FS, Smirniotopolous JG. Educational efficacy of computer-assisted instruction with interactive videodisc in radiology. Investigative Radiology 1993;28:1052-1058.
6. Okada RD, Kirshenbaum HD, Kushner FG, et al. Observer variance in the qualitative evaluation of left ventricular wall motion and the quantitation of left ventricular ejection fraction using rest and exercise multigated blood pool imaging. Circulation 1980;61(1):128-36.
7. Upton MT, Rerych SK, Newman GE, Bounous EJ, Jones RH. The reproducibility of radionuclide angiographic measurements of left ventricular function in normal subjects at rest and during exercise. Circulation 1980;62(1):126-32.
8. Rich S, Sheikh A, Gallastegui J, Kondos GT, Mason T, Lam W. Determination of left ventricular ejection fraction by visual estimation during real-time two-dimensional echocardiography. Am Heart J 1982;104(3):603-6.
9. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. N Engl J Med 1994;331(22):1493-9.